DOCUMENT RESUME

ED 287 431 HE 020 900

| | |
|---|---|
| AUTHOR | Alexander, Joanne M.; Stark, Joan S. |
| TITLE | Focusing on Student Academic Outcomes. A Working Paper. |
| INSTITUTION | National Center for Research to Improve Postsecondary Teaching and Learning, Ann Arbor, MI. |
| SPONS AGENCY | Office of Educational Research and Improvement (ED), Washington, DC. |
| PUB DATE | 86 |
| GRANT | OERI-86-0010 |
| NOTE | 43p.; From the NCRIPTAL Program on Research Leadership, Design and Integration. |
| AVAILABLE FROM | NCRIPTAL, Suite 2400 School of Education Building, The University of Michigan, Ann Arbor, MI 48109-2748 (Technical Report No. 86-A-002.0, $5.00). |
| PUB TYPE | Reports - Descriptive (141) -- Information Analyses (070) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | Academic Achievement; Basic Skills; *Classification; *College Students; *Educational Assessment; *Evaluation Methods; Higher Education; *Outcomes of Education; *Research Design; Standardized Tests; Student Development |
| IDENTIFIERS | *College Outcomes Assessment |

ABSTRACT

Current methods and instruments for assessing college student academic outcomes are identified and described, and possible outcome measures of NCRIPTAL's (National Center for Research to Improve Postsecondary Teaching and Learning) are suggested. Section I defines college outcomes from several perspectives, including pressures for outcome assessment, emphasis on quality, and issues that hinder assessment. Section II reviews various approaches to outcome assessment as well as existing typologies for classifying outcomes, including works of Pace, Astin, Bowen, Ewell, and Lenning. A typology tentatively adopted by NCRIPTAL researchers is introduced and delimits the discussions of outcomes in Section III, which presents several common outcome measures. They are grouped in three categories: (1) academic-cognitive outcomes (Graduate Record Examination, Undergraduate Assessment Program of the College Entrance Examination Board, American College Testing (ACT) Program achievement tests, ACT College Outcome Measures Project, College Level Examination Program, critical thinking and higher level outcome measures, basic skills); (2) academic-motivational outcomes; and (3) academic-behavioral outcomes (career/life goal exploration, diversity, persistence, faculty-student relationships). Appended are charts showing the categories of the NCHEMS Outcome Structures, and lists of basic skills tests for college students. Seventy references are included. (LB)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

NATIONAL CENTER FOR
RESEARCH TO IMPROVE
POSTSECONDARY
TEACHING AND LEARNING

ED287431

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCRIPTAL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

# Focusing on Student Academic Outcomes

A Working Paper

**Papers in this series available from** NCRIPTAL

**Approaches to Research on the Improvement of Postsecondary Teaching and Learning: A Working Paper**
*Patricia J. Green and Joan S. Stark*

**Postsecondary Teaching and Learning Issues in Search of Researchers: A Working Paper**
*Carol D. Vogel and Joan S. Stark*

**Teaching and Learning in the College Classroom: A Review of the Research Literature**
*Wilbert J. McKeachie, Paul R. Pintrich, Yi-Guang Lin, and David A. F. Smith*

**Psychological Models of the Impact of College on Students**
*Harold A. Korn*

**Designing the Learning Plan: A Review of Research and Theory Related to College Curricula**
*Joan S. Stark and Malcolm A. Lowther, with assistance from Sally Smith*

**Faculty as a Key Resource: A Review of the Research Literature**
*Robert T. Blackburn, Janet H. Lawrence, Steven Ross, Virginia Polk Okoloko, Jeffery P. Bieber, Rosalie Meiland, and Terry Street*

**The Organizational Context for Teaching and Learning: A Review of the Research Literature**
*Marvin W. Peterson, Kim S. Cameron, Lisa A. Mets, Philip Jones, and Deborah Ettington*

**Electronic Information: Literacy Skills for a Computer Age**
*Jerome Johnston*

**Design in Context: A Conceptual Framework for the Study of Computer Software in Higher Education**
*Robert Kozma and Robert Bangert-Drowns*

# Focusing on Student Academic Outcomes

## A Working Paper

by
Joanne M. Alexander
and
Joan S. Stark

Grant Number OERI-86-0010

Joan S. Stark, Director
Wilbert J. McKeachie, Associate Director

Suite 2400 School of Education Building
The University of Michigan
Ann Arbor, Michigan 48109-2748

(313) 936-2748

©1986 by the Regents of The University of Michigan

The project presented, or reported herein, was performed pursuant to a grant from the Office of Educational Research and Improvement/Department of Education (OERI/ED). However, the opinions expressed herein do not necessarily reflect the position or policy of the OERI/ED and no official endorsement by the OERI/ED should be inferred.

# Contents


Introduction 1

I. Outcomes and Outcome Assessment 3
- Current Pressures for Outcome Assessment 3
- An Emphasis on Quality 3
- Issues that Hinder Assessment 4
- Purposes of Outcome Assessment 5
- Locus of Assessment 5
- NCRIPTAL's Research Mission and Model 6

II. Approaches to Outcomes and Outcome Assessment 9
- NCRIPTAL's Delimited Outcome Framework 12
- Type of Outcome Measures 12
- Level of Outcome Measures 14
- Form of Outcome Measures 14

III. Outcome Measures and Outcome Research 17
- Academic-Cognitive Outcomes 17
  - Graduate Record Examination 17
  - American College Testing Program Achievement Tests 19
  - Undergraduate Assessment Program of the College Entrance Examination Board: Area Tests and Field Tests 19
  - ACT College Outcome Measures Project 20
  - College Level Examination Program 20
  - National Teacher Examination 20
  - Critical Thinking and Higher Level Outcome Measures 21
  - Basic Skills 23
- Academic-Motivational Outcomes 23
  - Involvement and Quality of Effort 23
- Academic-Behavioral Outcomes 24
  - Career and Life Goal Exploration 24
  - Exploration of Diversity 25
  - Persistence 25
  - Faculty-Student Relationships 25
- Conclusion 26

Appendixes 27

A. Categories of the NCHEMS Outcome Structures 27
B. Basic Skills Tests for College Students 31

References 35

## Tables


1. A Typology of Student Outcomes (Astin, Panos, & Creager) 10
2. Outcomes Over Time (Astin) 10
3. Propositions and Caveats About Type, Level, and Form of Outcome Measurement 13
4. A Whole-Person Approach to College Student Outcomes 13
5. NCRIPTAL's Outcome Framework 14


## Figures


1. Parameters of Outcome Assessment 6
2. Variables in NCRIPTAL's Research Agenda 7

# Introduction

The purpose of this working paper is (1) to identify and describe some current methods and instruments for assessing college student academic outcomes, (2) to suggest possible outcome measures of NCRIPTAL's research program, and (3) to suggest methods of outcome assessment for other researchers and practitioners. In the process of this exploration we review existing literature on outcome assessment and other related research literature on college outcomes to set the stage for a discussion of appropriate methods for recognizing improved teaching and learning. We also examine typologies and frameworks for understanding outcomes developed by several scholars.

The literature on college outcomes and their measurement is evolving rapidly. We acknowledge a substantial debt to several scholars, particularly Alexander Astin, Howard R. Bowen, Peter Ewell, C. Robert Pace, Ernest Pascarella, and others, from whose earlier reviews we have selected and summarized material liberally. Along with these scholars and other organizations, such as the American Association of Higher Education, which has gathered a substantial collection of literature on outcome assessment, NCRIPTAL desires to be of service to educators who seek a broad, nontechnical summary of this emerging field.

Undoubtedly, there is new and important literature that we have overlooked or that is still in press. It is particularly difficult to know of valuable work in progress at individual campuses but we believe that this general summary will encourage practitioners and researchers to inform us of their successes as well as the difficulties they encounter. Therefore, this paper is a working document to be updated periodically as NCRIPTAL learns of the work of educators who have developed new measures and new techniques.

This paper is a working document in another sense as well. It has been developed during the first few months of our existence as a national center to conduct research and provide leadership in improving postsecondary teaching and learning. Concurrently, several NCRIPTAL researchers are developing reviews of outcome measures that can be used to assess specific aspects of student academic development. Because of their concurrent development and the technical nature of these related reviews, they are mentioned only briefly in this overview. A list of the concurrently developed papers follows the title page.

As NCRIPTAL's work proceeds over the next several years, an important goal is to understand connections among more specific areas of research on student growth. Thus, future syntheses will describe and develop more completely the relationships among potential measures of student outcomes within the NCRIPTAL typology described later in this paper.

The paper is organized in the following manner: Section I defines college outcomes from several perspectives and discusses the importance of outcome assessment at the postsecondary level. Pressure from the academic community and from federal and state agencies has increased interest among educators in assessing student achievement. The resulting interest has brought into focus a number of issues about outcome assessment, including choices among appropriate models for outcome assessment. NCRIPTAL's mission in improving postsecondary teaching and learning and its relation to outcome assessment is introduced.

Section II reviews various approaches to outcome assessment as well as existing typologies for classifying outcomes. The recent works of Pace, Astin, Bowen, Ewell, Lenning, and others are discussed. A typology tentatively adopted by NCRIPTAL researchers is introduced and delimits the discussions of outcomes in Section III.

Section III presents a number of common outcome measures as potential measures both for NCRIPTAL and for the research community and educators in general. Areas of the NCRIPTAL typology in which outcome measures are underdeveloped are noted.

# I. Outcomes and Outcome Assessment

Before engaging the issues of outcome assessment, a review of the definitions of outcomes that have been used by researchers is in order. In their groundbreaking review, Feldman and Newcomb (1969) refer to the *impact* of college on students rather than to *outcomes*. They view impact as the influence of colleges on student orientations and characteristics. Bowen (1977) takes an economic approach to outcomes, defining them as the result of transformed institutional resources. The primary product of the transformation is individual learning; additional products include changes in other intangible individual qualities.

Pace (1979) defines outcomes as changes that are widely accepted as goals of higher education and that are the result of events and experiences in college designed to help students attain these goals. Astin (1980) uses a value-added approach to outcomes. He specifies that outcomes are the measured differences between entry characteristics of a student and the characteristics of a student on exit from college. Ewell's (1983) definition of outcomes coincides with Astin's. He defines outcomes as any change or consequence that occurs as a result of enrollment in an educational institution and participation in its programs.

The major differences among these definitions relate to whether they address the question of what the outcomes are or the question of why outcomes occur. Pace (1979) claims that Feldman and Newcomb, through the use of the term *impact*, are attempting to explain the causes of certain outcomes. The question that Feldman and Newcomb address, therefore, is why certain changes occur. In contrast, Pace prefers simply to address the issue of change. He claims that by measuring change the "what" question being addressed is: What are the outcomes of college? This question, he believes, is much simpler to address and provides primary evidence of the results of a college education.

For the purpose of NCRIPTAL's research program for improving teaching and learning, both the "what" and the "why" questions are important. This paper focuses on determining what changes occur; other NCRIPTAL literature reviews focus on various reasons for the outcomes. This paper also addresses the question of which measures may be most appropriate for measuring the outcomes of postsecondary learning.

The outcome definition most closely suited to NCRIPTAL's model is Astin's definition of outcomes. NCRIPTAL's use of a value-added, change model will assist in the discovery of the effects of various instructional, programmatic, and individual characteristics on the teaching and learning process. The framework NCRIPTAL has adopted to focus its work is presented in Figure 1 (see page 7).

### Current Pressures for Outcome Assessment

Currently there are pressures in postsecondary education from two directions for outcome assessment—from the academic community itself and from employers of college graduates. First, the academic community is calling on postsecondary institutions to use assessment as a means of improving the quality of education. A number of books and articles have pointed to an apparent lack of quality control in collegiate education and have suggested measurement of student progress as one means of rectifying the situation. The National Institute of Education's *Involvement in Learning* (NIE Study Group, 1984) calls for the systematic assessment of students' knowledge, capacities, and skills as a way of addressing problems in the undergraduate curriculum. In *To Reclaim a Legacy* (1984), William Bennett also called for curricular reform, minimum standards, and assessment as a way of standardizing the meaning of the undergraduate degree. *Integrity in the College Curriculum* (Association of American Colleges, 1985) posits that the absence of institutional accountability is a grave problem and that the measurement of student progress poses a solution to the dilemma. *Access to Quality Undergraduate Education* (Southern Regional Education Board, 1985) calls for a cooperative effort within the educational community to find ways of improving quality while maintaining access. The report suggests that new ways of measuring student progress and performance are needed to resolve the access/quality problem.

### An Emphasis on Quality

A key issue raised in all of these reports is quality. For years the worth of a college education went unquestioned; it was typically assumed that college graduates left college with an increased amount of knowledge and understanding.

Today, however, public attitudes have changed. A number of indicators point to a decrease in the quality of college education and this has resulted in a call for accountability for improved teaching and learning. Among the cited indicators are: (1) a large number of students who need remedial courses at the college level, (2) a decline in student scores on verbal sections of standardized tests, (3) a decline in graduate scores on standardized tests and professional licensing exams, and (4) an increased number of students pursuing professional and occupational studies rather than a liberal arts education (Hartle, 1985). Despite disagreements over appropriateness of these indicators, many believe the quality problem is real.

These reports have pointed to assessment as a solution to the problem. The common themes in these reports include a need for stronger student performance, clearer expectations of what college students should learn, and more rigorous measurement of educational achievement.

The expressions of concern about educational quality are not unique to the educational/academic community. Spokespersons from private industry, government, and accreditation agencies have called for increased institutional accountability as well (Ewell, 1985). The private sector, as the major employer of college graduates, desires more uniformly high quality in the graduates it hires. In addition, state legislators are becoming increasingly concerned with the return states are getting from their investment in higher education. These constituencies represent the second pressure calling for postsecondary assessment.

**Issues that Hinder Assessment**

Despite this increased pressure on institutions to evaluate and assess student outcomes, very few colleges actually have established student assessment programs (Ewell, 1985). This limited response may be due to a number of concerns and problems revolving around student assessment.

One of the most difficult issues surrounding outcome assessment is the question of what the outcomes of college should be. At the secondary level, the assessment of basic skills provides a commonly accepted level of achievement. At the college level, however, there is no common base level of higher academic skills that is universally accepted. As Turnbull (1985) stated,

> Beyond basic skills there lies an immense realm of disagreement about collegiate goals....It is essential to realize that the purposes of higher education are a matter of fundamental debate. (p. 24)

This lack of consensus exists within colleges as well as between colleges. Hartle (1985) recognizes the problem within institutions.

> The central problem is that measuring educational achievement may well require more agreement about the ends and means of a higher education than we have at most institutions. (p. 15)

The development of consensus on minimum requirements is crucial for developing a successful outcome assessment program within a given institution. Consensus among institutions would be even more difficult to attain and might result only in agreement on very minimal outcomes.

A common set of stated outcomes for all institutions would make it difficult to take into account the broad range of institutional goals and missions. To be useful and effective, assessment programs must address the diversity of institutional goals. In a society with a diversified and decentralized system of postsecondary education, assessment programs must be tailored to fit the needs of the various institutions.

In his discussion of accountability, Bowen (1974) recognizes the diverse goals of institutions and calls for a matching of assessment programs to college goals. He proposes that to attain true institutional accountability an institution must (1) define goals and order priorities, (2) measure and identify outcomes, (3) compare the outcomes with the goals to determine the degree to which goals have been met, and (4) measure the cost and determine whether it is reasonable (p. 2).

Once the question of consensus and outcomes has been resolved, the method of measurement arises as an issue. Numerous measures are available for measuring both cognitive and affective outcomes among students. Deciding on the appropriate measures is difficult. Harris (1985) offers guidelines for practitioners getting started with an assessment program in higher education.

Assessing improved teaching and learning is also somewhat hindered because of political problems it can pose for faculty and administrators. Ewell (1983) identifies a number of concerns that cause administrators to avoid assessment. First of all, they fear that no positive impact will be found and that results will reflect badly on their leadership. Ewell believes this fear is unfounded, given a number of successful attempts in finding positive outcomes. Administrators are also concerned about the misinterpretation of quantitative results. Although some outcomes may be qualitative in nature, the measures used most typically are quantitative descriptors. Administrators and faculty fear that people may place too much weight on the numbers, losing sight of the fact that the measures

are only proxies for the actual outcomes. In addition, the false precision of quantitative measures may further complicate the public interpretation of the results. These fears among educators have hindered the development of student outcome assessment programs.

**Purposes of Outcome Assessment**

Even if agreement could be reached on what the outcomes of college should be, assessment holds varied meanings for various parties. According to Hartle (1985), the term assessment, in education, is often used interchangeably with evaluation and measurement. Yet, there are subtle differences among these terms. Assessment is the process of gathering data (measurement) and assembling evidence into an interpretable form for some intended use. Once the information is gathered, judgments (evaluations) may be made based on the evidence. Measurement, therefore, is only a part of the assessment process, and is not necessarily synonymous with assessment, whereas evaluation implies judgments based on the collated measures.

In addition to the problem of defining assessment, individuals tend to associate assessment with numerous and diverse activities. Hartle (1985) mentions six separate and non-parallel but overlapping activities that may be believed to represent assessment activities in higher education. One activity uses multiple measures and observers to monitor students' intellectual and personal growth. Another assessment activity may be associated with state-mandated requirements for evaluating student progress or academic program success. A third assessment activity is the value-added method for measuring student progress, which involves pre- and post-testing and attribution of gains to the college experience. A fourth activity involves the use of standardized testing to measure the extent of student knowledge. A fifth assessment activity uses assessment in fund allocation, frequently by rewarding institutions based on performance criteria. Finally, measurement of changes in student attitudes and values is a sixth activity considered part of the assessment domain.

**Locus of Assessment**

In addition to the issues surrounding agreement on assessment outcomes and activities, issues arise over the level of analysis at which assessment occurs. Data gathering for assessment programs can occur at one or more of three levels: the individual student, the academic or department program, and the institution. The usefulness of assessment at each level depends on the purpose of the investigation. Some individuals have argued that assessment should focus on the individual learner (Hartle, 1985; Bowen, 1979). Others view the academic department as the appropriate level of analysis for investigating teaching and learning environments (Winteler, 1981). At a higher level of aggregation, measures have been developed to assess institutional outcomes. Pascarella (1985), however, asserts that institutional differences are much more difficult to pinpoint and that differences in outcomes are more effectively explained through individual characteristics.

One final important issue concerning outcome assessment is whether the process is internally or externally administered. While some institutions have established their own programs (e.g., Alverno College), others have received state mandates to use standardized testing to assess the quality of education (e.g., Tennessee and Florida). In general, academic institutions tend to feel threatened by external control and fear loss of autonomy. Some assessment proponents indicate that these fears should inspire institutions to initiate internally directed programs before state level initiatives are realized. Institutional assessment programs may have needed flexibility to address most appropriately the particular goals of the institution.

There is strong opinion among some authorities that the pressure for educational assessment will not dissipate. For example, basing his conclusion on several factors, Hartle (1985) warns that assessment is not a passing fad. First, now that the issue of student access has been resolved, the focus on quality assurance is essential. Second, there is a widespread public concern over the lack of value placed on teaching in some colleges. Finally, state governments are well-informed and interested in the quality of education.

Such predictions, and the abundance of existing definitions and interpretations about them, indicate a continuing need to specify at least eight parameters for effective discussion about assessment of student outcomes.

1. What are the *purposes* or *incentives* for assessment? Assessment activities to satisfy state mandates may differ substantially from those undertaken to improve learning within an institution.

2. The *type* of assessment being discussed is important. For example, is information on student values, student academic achievement, or the employment rate of new graduates to be gathered? Is outcome information to be gathered only on those outcomes upon which some group has achieved consensus or, with the possibility of improving under-

standing or consensus, should it also be gathered in areas about which little agreement exists?

3. At what *level* is the assessment to occur? For example, is it important to learn about progress of individual students or groups of students, about students in specific academic programs, about the programs themselves, or about the institution as an aggregate of students and programs?

4. What will be the *form* of assessment? For example, is the assessment to focus on measures of student change (the "value-added" approach) or to determine whether students have reached some expected level of achievement? Can these two forms be effectively combined? Will multiple measures or a single measure of each outcome be used?

5. What agency will be responsible for *administration* of an assessment program? Possibilities range from external groups such as state agencies and legislatures to groups of faculty in specific academic programs.

6. If evaluation is to follow assessment, what will be the *locus of evaluation*? That is, once the information has been gathered in an interpretable form, who will make evaluative judgments about what types of change or stability are suggested?

7. What will be the *locus of decisions* about the appropriate use of the information or about any evaluative judgments that are made?

8. What will be the *use* of the evaluative judgments made on the basis of assessment activities? For example, if judgments are made, will they be about the merit or worth of students, the merit or worth of programs, the merit or worth of institutions? Or will they be focused on specific recommendations for improvement of student learning, of program quality, or institutional functioning?

These eight parameters may be interrelated in many ways; certainly decisions about each will influence decisions about the others. Many observers would agree, however, that the most crucial linkage is that between the *purpose* of assessment and *use* of data gathered in the assessment process. Another logical linkage (perhaps less clear) may exist between the level at which assessment is undertaken and the level of administration. Leaving other linkages unspecified as unique to a given situation, we have illustrated the importance of these relationships by arranging the eight parameters as shown in Figure 1.

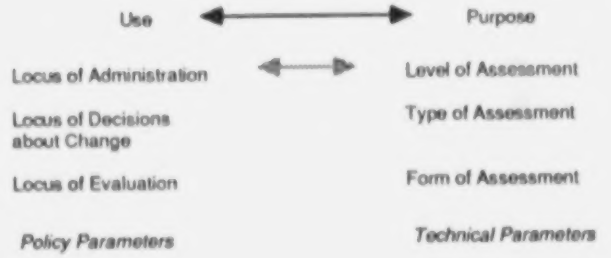



Figure 1. Parameters of Outcome Assessment

One implication of the arrangement in Figure 1 is that the parameters on the left, those having to do with loci of assessment and subsequent evaluations and decisions are basically policy issues. Those on the right, describing type, form and level of assessment, are largely technical issues. Both sets strongly depend on the purposes and planned uses of assessment. In the following section, we describe briefly NCRIPTAL's mission. Our purpose in doing so is to show how our efforts assume a specific institutional purpose for assessment activities, namely, improvement in the teaching and learning environment. We expect, therefore, that institutions will also be responsible for using information they gather through use of outcome measures. Consequently, the type, level, and form of outcome measures selected, developed, and used in NCRIPTAL's research agenda will be those most suitable for use by institutions interested in improving teaching and learning.

**NCRIPTAL's Research Mission and Model**

The National Center for Research to Improve Postsecondary Learning and Teaching will focus its research, development, and dissemination activities on five aspects of college learning environments that affect learner outcomes: classroom learning and teaching strategies, curricular structure and integration, faculty attitudes and teaching behaviors, organizational practices, and the use of emerging information technology. While recognizing multiple student outcomes of college, such as cognitive development, personal development, and career development, the Center initially will emphasize cognitive development of undergraduate students in colleges that concentrate on teaching as their primary mission. This emphasis was chosen because the recent dramatic progress of research in cognition holds great promise for improving learning and teach-

ing. Furthermore, student cognitive development is intimately linked to career development and to other important outcomes such as the development of a sense of self-efficacy, personal responsibility, and motivation.

Student's cognitive and affective characteristics, which vary with their diverse backgrounds, are important conditioners as well as predictors of learning experiences. Since learners of many backgrounds and ages now attend college and since instructors may select an increasing variety of potentially effective strategies, the Center will attempt to discover optimum combinations of learner characteristics and instructional processes to facilitate cognitive development.

To complete this mission, a research framework for the NCRIPTAL's work has been developed. Considered simply, this model (see Figure 2) includes three general research variables: student characteristics (independent variables), teaching/learning environments (alterable variables), and student outcomes (dependent variables). Student characteristics are motives, learning styles, prior knowledge, skills, and other characteristics that students bring with them to college. These characteristics interact with institutional environments to determine learning.

The teaching/learning environments are influenced by the faculty, the curriculum, the teaching and learning strategies, the institutional practices, and by the technological environment. Although extracurricular and interpersonal factors also influence the institutional environment, these variables will not be included in NCRIPTAL's research agenda. Additionally, the environmental factors interact and overlap, and these interactions will be recognized in our research agenda.

Student outcomes here are defined as the results of students' involvement in teaching/learning environments. These outcomes may be both long-term (measurable throughout life after completion of college) and short-term (measurable during or immediately following the college experience), but NCRIPTAL will focus on the short-term outcomes as most directly applicable to improving teaching and learning. Although briefly characterized in the usual manner as "independent" variables, student characteristics and goals may also change as a result of the college experience. Thus, a feedback loop is included in the Center model indicating that education is an iterative process whereby the typically independent variables are affected over time by the teaching/learning environments.

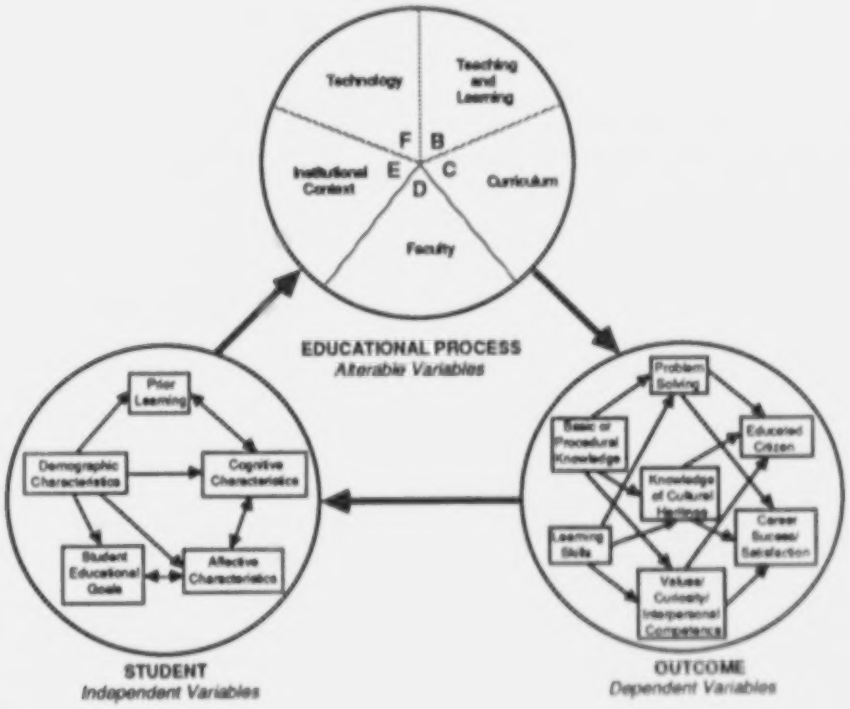



Figure 2. Variables in NCRIPTAL's Research Agenda

# II. Approaches to Outcomes and Outcome Assessment

A number of researchers have attempted to classify outcomes and specify approaches for assessing outcomes. In this section, approaches by Ewell, Astin, Lenning et al., and Bowen are discussed.

Ewell (1983) discusses three approaches that have been used to measure student outcomes: academic investigation perspective, student-personnel perspective, and management perspective. Actually these approaches are based on the purpose of the investigators and thus use different perspectives on outcomes, have different goals for using outcomes, and involve different data requirements.

Academic investigation (research) is the oldest and most commonly used reason for measuring student outcomes. The college experience is investigated in a typical research fashion: theories about student growth are developed, tested, and refined as a result of data collection. From this perspective, most of the research on student outcomes has been done by psychologists and sociologists. Frequently psychologists have focused on the impact of college on personal and cognitive development and sociologists have concentrated on such issues as the impact of college on social mobility and socialization of students into the professional fields. In this perspective the goal to explain (and ultimately to predict) human behavior and the data collected must have high empirical quality and be objective. While some of the relationships discovered in this research have been used by institutional policymakers, it should be noted that decisional utility is not the goal; the purpose is to successfully account for a given outcome.

The student personnel approach uses student outcomes as a means for evaluating students for admission to programs and placement on completion of the program. The data are also used for counselling students in career selection and for evaluating the effectiveness of programs for meeting student needs. In this perspective the goal of outcome measurement is to gain assessment information about individual students. Data is considered useful if it provides information for student placement or if it is diagnostic of student problems. The theoretical constraints of data collection are not crucial when using this approach.

The management perspective for measuring outcomes is a still different approach to outcome assessment. From this perspective the focus is on the use of outcome assessment as a method to improve administrative decisions, particularly those involving program planning and budgeting. The goal of outcome assessment in this perspective is to improve the quality of resource-allocation decisions. To meet this goal, data must be empirically valid, reliable, and perceived by the decision makers as relevant to the decision.

Ewell's classification of approaches to student outcomes is useful because it calls attention to varied uses of outcomes and the ways in which different goals influence the collection of student outcome information.

In addition to classifying approaches to outcome assessment based on proposed uses, researchers have attempted to classify types of educational outcomes. Astin (1974) developed a taxonomy of student outcomes involving three dimensions: type of outcome, type of data, and time. The types of outcome are split into two domains: cognitive and affective. The cognitive domain includes outcomes such as basic skills, general intelligence, and higher-order cognitive processes. The affective domain includes outcomes often described as attitudes, values, and self-concept.

The data dimension is also split into two domains: behavioral and psychological. This dimension distinguishes between outcome data that are covert and those that are observable. The behavioral domain refers to observable activities of the individual. The psychological domain refers to the internal states or traits of the individual. While the actual outcomes may be the same, the ways in which the information is gathered to represent them are different.

The primary two dimensions of Astin's approach are shown in Table 1. This typology has been widely accepted as a method for classifying outcomes. In Astin's typology the third dimension, time, stresses the importance of including both the long- and short-term outcomes of college. Some examples of applying the time dimension to the outcome cells are provided in Table 2.

In addition to the typology, Astin (1974) provided some insights into the assessment of educational outcomes. To him the fundamental purpose of assessment is to produce information that is useful for decision making. Thus measurement should begin with a value statement—an idea about what future state would be desirable or important.

Lenning and Associates (1983) at the National Center for Higher Education Management

**TABLE 1**

**A Typology of Student Outcomes**

| DATA | OUTCOME | |
|---|---|---|
| | Affective | Cognitive |
| *Psychological* | Self-concept<br>Values<br>Attitudes<br>Beliefs<br>Drive for Achievement<br>Satisfaction with College | Knowledge<br>Critical Thinking Ability<br>Basic Skills<br>Special Aptitudes<br>Academic Achievement |
| *Behavioral* | Personal Habits<br>Avocations<br>Mental Health<br>Citizenship<br>Interpersonal Relations | Career Development<br>Level of Educational Attainment<br>Vocational Achievements<br>Level of Responsibility<br>Income<br>Awards or Special Recognition |

Source: Alexander W. Astin, R.J. Panos, and J.A. Creager, *National Norms for Entering College Freshmen - Fall 1966* (Washington, D.C.: American Council on Education, 1967): p. 16.

**TABLE 2**

**Outcomes Over Time**

| OUTCOME | DATA | SHORT-TERM INDICATOR | LONG-TERM INDICATOR |
|---|---|---|---|
| Affective | Behavioral | Choice of major field of study | Current Occupation |
| Affective | Psychological | Satisfaction with college | Job Satisfaction |
| Cognitive | Behavioral | Persistance | Job Stability |
| Cognitive | Psychological | LSAT score | Score on law boards |

Source: Astin, 1974, p. 33

Systems (NCHEMS) developed an extensive framework for identifying the universe of major "outputs" and outcomes of postsecondary institutions. In developing this taxonomy, the authors sought to develop an exhaustive list of outcomes to assist in the assessment of managerial effectiveness. As a result of the management perspective, Lenning et al. did not focus exclusively on student outcomes but rather included them in two of the several categories: human characteristics outcomes and knowledge, technology, and art forms outcomes. Viewed in Astin's terms, the human characteristics outcomes include primarily affective and personality characteristics, as well as skill outcomes. The knowledge, technology, and art form category includes the typically cognitive outcomes: both specialized and general knowledge and scholarship. Additional outcome categories in this framework include (1) economic (e.g., economic security, standard of living), (2) resource and service provision (e.g., teaching, facility provisions), and (3) other maintenance and change (e.g., traditions, organizational operation). A listing of the complete NCHEMS taxonomy is included in Appendix A. Clearly this framework includes both long- and short-range student outcomes as well as outcomes at the program and institutional level.

Bowen (1974) took a slightly different approach from the two previous researchers when discussing outcomes. Instruction is related to the outcome of learning and changes in human traits. Research and scholarship relates to the outcomes of preservation, discovery, and interpretation of knowledge, artistic and social criticism, philosophical reflection, and advancement of the fine

arts. Public service results in societal outcomes such as improved health, solutions to social problems and agricultural productivity (p. 2-3).

Of these three services, Bowen believes that instruction is the primary goal of higher education and bringing about desired changes in students is central to this mission. Bowen's approach could be viewed, therefore, as primarily academic in nature. He focused on investigating the changes that occur among students without emphasizing the use of these measures in either placement or decision making.

In a later work, Bowen (1977) broadened his view of student learning and offered a more elaborate catalogue of accepted goals. This catalogue of goals serves also as a typology of student outcomes derived from three widely accepted goals of instruction. These three general goals are: educating the whole person, addressing the individuality of students, and maintaining accessibility. The first goal, educating the whole person, refers to the idea that education should cultivate both the intellectual and affective dispositions of persons, thereby enhancing intellectual, moral, and emotional growth. The second goal, addressing individuality, requires that the uniqueness of individuals be taken into account in the educational process. Accessibility refers to the notion that education should be readily available to a broad range of persons.

According to Bowen, the catalogue of goals derived from these general goals constitutes both a model for the educational system and the criteria by which the system can be judged. While Bowen recognized that his goal typology has utopian qualities, he posits that it provides a useful model that can be used to shape and guide institutional functioning.

In Bowen's scheme specific educational goals are divided into two groups: goals for individual students and goals for society. The five categories of goals for individual students include: cognitive learning, emotional and moral development, practical competence, direct satisfactions from college education, and avoidance of negative outcomes. In a further subdivision, cognitive learning includes ten specific areas of learning. They are:

1. Verbal skills: Ability to read, speak, and write clearly and correctly.

2. Quantitative skills: Understanding of mathematical and statistical concepts.

3. Substantive knowledge: Acquaintance with Western culture and traditions and familiarity with other cultures. Knowledge of contemporary philosophy, art, literature, natural science, and social issues. Understanding of facts, principles and vocabulary within at least one selected field.

4. Rationality: Ability to think logically, and analytically, and to see facts clearly and objectively.

5. Intellectual tolerance: Openness to new ideas, curiosity, and ability to deal with ambiguity and complexity.

6. Esthetic sensibility: Knowledge of and interest in literature, the arts and natural beauty.

7. Creativeness: Ability to think imaginatively and originally.

8. Intellectual integrity: Respect for and understanding of the contingent nature of truth.

9. Wisdom: Ability to balance perspective, judgment and prudence.

10. Lifelong learning: Sustained interest in learning. (Bowen, 1977, pp. 35-36)

Bowen's remaining four categories of student goals are focused primarily on affective and long-term student outcomes.

Bowen also suggested seven principles that should be used in the identification of outcomes and thus in outcome assessment at particular colleges. The first principle is that inputs should not be confused with outputs. Bowen claims that high institutional expenditures (an input) do not guarantee equivalently high outcomes; the differences between inputs and outputs has too often been ignored. The only valid outcome measurement is of the development and changes that occur in students as a result of their college experience.

The second principle suggests that assessment should be linked to all educational goals, not just to those developments easily measured or related to economic success. Bowen offers his catalogue of goals as a starting point on which to build an assessment plan.

The third principle states simply that educational outcomes should relate to the person as a whole; and the fourth principle posits that outcome assessment should include the study of alumni as well as current students. The fifth principle suggests that outcome assessment should measure changes that occur as a result of the college experience.

The sixth principle states that an evaluation scheme must be practical: not too time-consuming or expensive. The assessment should focus

on major goals of the institution and need not be based on the entire population of students. However, results must be reported in a form that the general public can read and understand.

The final principle asserts that assessment should be controlled from within the institution rather than being imposed by external agencies. Assessment programs should be designed for each institution, keeping the special missions and philosophies of the institutions in mind.

Ewell (1983) mentions additional outcome dimensions that should be considered. These include whether (1) the effects are short- or long-term, (2) the student is aware or unaware of the outcome, (3) the effect is direct or indirect (i.e., how closely the outcome is connected to the educational program), and (4) the outcome is intended or unintended. These dimensions represent important differences between outcomes that should be considered in outcome research and assessment.

In more recent work than that reviewed earlier, Astin (1979) identifies three core measures of student outcomes that should be included in a student outcome data base. First, students' successful completion of a program of study should be included. More specifically, information is needed to determine whether students' accomplishments are consistent with their original goals. Second, a measure of cognitive development must be included and more than grade point average and class standing are needed. Preferably, repeated measurement will be used so that change can be assessed by comparing performance at two points in time. Third, measures of student satisfaction should include satisfaction with the quality of the curriculum, teaching, student services, facilities, and other aspects of the college.

Beyond these essential measures, the student data should include information gathered on entry, during the educational process, and at exit or another designated point of time. Student characteristics should be recorded when they first enroll, information on what happens to the student while enrolled at the college must be available, and measures of the degree of attainment of desired or behavioral objective at exit must also be accessible. This approach, developed by Astin, is known as the "value-added" approach. It asserts that outcome measures alone tell us very little about institutional effectiveness or impact. By controlling for entry characteristics, however, a more accurate picture of outcomes will emerge. In the absence of such data, outcome measures may be grossly misinterpreted when used for assessing institutional effectiveness because most outcomes are highly dependent on the characteristics of students at entry.

**NCRIPTAL's Delimited Outcome Framework**

As discussed earlier, NCRIPTAL's mission includes both conducting basic research on the effects of various aspects of the teaching and learning environment on student outcomes and providing leadership and assistance to institutions in their own assessment and evaluation efforts. Thus, in the terms of Ewell's "perspectives," we must engage in a dual approach, combining the academic-investigative spirit of basic research and a management perspective that can help institutions construct their own assessment processes and uses of the information.

Fulfilling this dual mission with available resources requires delimitation of the arena in which our work will be conducted and a selection of outcome measures and assessment principles that seem most closely related to practical concerns in improving teaching and learning. Existing typologies, such as that proposed by Astin (see Tables 1 and 2), the list of principles by Bowen, and the important distinctions mentioned by Ewell, as well as the work of many other scholars, have been helpful in formulating our plans. In Table 3 we have summarized some of these propositions, attempting to group them as accurately as possible under the "technical parameter" headings discussed earlier, namely, "type of outcome to be measured," "level of measurement," and "form of measurement." This grouping forms the basis for our discussion of outcome measures to be used in NCRIPTAL's work. It bears repeating that only these three parameters of type, level, and form are discussed because we have already focused our work on a specific purpose (improvement of teaching and learning) and assume that results will be used for decisions consonant with that purpose. Furthermore, our efforts are based on the assumption that the administrative locus of assessment activities and evaluative decisions about this information all rest within the college or university.

**Type of Outcome Measures**

However desirable it might be for researchers and institutions to follow Bowen's suggestion to assess all possible outcomes and relate outcomes to the development of the whole person, such a global program would readily encounter problems of feasibility and lack of consensus. Nonetheless, our discussion of outcome measures begins with the whole-person approach in an effort to determine which subsets of this universe are of greatest importance.

During such discussions we found many benefits, but some pitfalls, in Astin's encompassing four-fold typology of student outcomes (see

**TABLE 3**

**Propositions and Caveats about Type, Level, and Form of Outcome Measurement**

| | BOWEN | EWELL | ASTIN |
|---|---|---|---|
| *Type of Outcome Measures* | Assess all outcomes, even those difficult to measure. | Distinguish intended and unintended outcomes. | |
| | Relate outcomes to whole person. | Distinguish outcomes of which student is aware and unaware. | |
| | Focus on changes attributable to college. | Distinguish outcomes closely linked to educational program. | |
| | Focus on major institutional goals. | | |
| *Level of Outcome Measures* | Study alumni as well as current students. | Distinguish short- and long-term outcome measures. | Record whether students completed program and whether accomplishments were consistent with their goals. |
| *Form of Outcome Measures* | Separate inputs and outputs. | | Measure at various points in time; include information at entry, during program, and on exit. |
| | Use practical and feasible means. | | Use measures of cognitive development beyond grade point average. |
| | | | Include measures of student satisfaction. |

Table 1). Specifically, although Astin acknowledged interactions between affective and cognitive outcomes, his typology used these concepts as two different primary dimensions. Consequently, the typology made little provision for attention to cognitive-personal outcomes or affective-academic outcomes. Yet, many cognitive psychologists and personality theorists believe that, particularly for students who enter college with undeveloped motivation or low self-efficacy, affective outcomes may be related to academic as well as to personal and social growth. As a result of these and related discussions, we drew a slightly different type of typology framework which notes three "arenas" of student growth in college and three forms through which changes in these arenas may be observed. The resulting nine-cell framework, which we stress was derived a priori from our accumulated experience, is shown in Table 4.

The *arena* dimension refers to the various aspects of life in which the outcome is important. The three arenas are personal, social, and academic. The personal domain includes outcomes like personal worth, feelings about oneself, satisfaction with personal accomplishments, ability to make decisions, and using one's skills appropriately. The social arena outcomes include ability to function in interpersonal relationships, citizenship, social responsibility, social awareness, and

**TABLE 4**

**A Whole-Person Approach to College Student Outcomes**

| *FORM OF DEMONSTRATED CHANGE* | ARENAS OF GROWTH AND DEVELOPMENT | | |
|---|---|---|---|
| | Social | Personal | Academic |
| *Cognitive* | | | |
| *Motivational* | | | |
| *Behavioral* | | | |

contributions to society. The academic arena includes academic achievement, self-efficacy, motivation, critical-thinking abilities, problem-solving skills, and goal exploration behaviors.

The form dimension also has three categories: cognitive, motivational, and behavioral. This dimension specifies the form in which the outcome is demonstrated. Cognitive outcomes are internal outcomes. Typically they occur within individuals' mental processes and their existence is inferred, usually through testing. Motivational outcomes consist largely of the feelings that individuals have about themselves, their capabilities, and the world around them. These outcomes are generally self-reported, though some social-psychological methods exist that tap these attitudes more discretely. Behavioral outcomes may be reported by the individual or directly observed.

As mentioned earlier, NCRIPTAL's research program will focus on the academic arena shown in Table 4. In selecting this subset of the universe of college outcome measures for attention, we risk posing for others the same difficulty that Astin's typology posed for us. We acknowledge that the personal and social arenas cannot be separated from the academic arena; one's personal and social development affects one's academic development and the reverse is also true. Nonetheless, by constructing a framework that includes three cells, academic-cognitive, academic-motivational, and academic-behavioral, we are able to encompass a broad set of outcomes of primary concern to colleges and the public as well as to incorporate recent theories of cognitive development. Table 5 shows a more detailed view of the academic arena and the types of outcomes that seem to fit into each of the three major cells.

At first glance, some observers will believe we have violated Bowen's principle of separating inputs and outputs by classifying as outcomes some of those items listed in the academic-motivational cell. Traditionally, motivation, self-efficacy, involvement, and effort have been viewed as fixed attributes students bring to the educational process. Our view that these characteristics are subject to change (in an intended or unintended direction) as a result of the educational process is, in part, what caused us to modify previously existing outcome typologies. Although little attention has been given to these ideas, most colleges would agree, for example, that improved motivation is an outcome to be sought. While the original motivation a student brings to college is an input, a new motivational level based on educational experiences becomes an outcome the student takes to the next stage of learning.

An additional previously neglected aspect of the iterative outcomes conception relates to Ewell's distinction between student awareness or lack of awareness of changes. Although we have not included it in the list at this time, if students are to take increased responsibility for their learning, awareness itself may be an outcome to be sought.

TABLE 5

**NCRIPTAL's Outcome Framework**

| *FORM OF MEASUREMENT* | ACADEMIC ARENA |
|---|---|
| *Cognitive* | Achievement (facts, principles, ideas, skills)<br>Critical-thinking skills<br>Problem-solving skills |
| *Motivational* | Satisfaction with college<br>Involvement/effort<br>Motivation<br>Self-efficacy |
| *Behavioral* | Career and life goal exploration<br>Exploration of diversity<br>Persistence<br>Relationships with faculty |

### Level of Outcome Measures

As already mentioned, both practicality and technical difficulties have caused us to set aside Bowen's suggestion that alumni be studied in addition to current students. Instead, NCRIPTAL's agenda will focus on outcome measures that can be related directly to classroom and program educational experiences. In general, our unit of analysis will be the individual student and groups of students sharing a common educational experience in a course or program. Whenever possible, outcome measures for special populations of students (e.g., minorities, women, adult students) will be examined in relation to similar data for traditional students.

Astin's point about whether students' eventual accomplishments are consistent with their goals will be a special focus of one of our research programs. In fact, goals of students at college entry are subject to change in both intended and unintended directions. Since there would likely be disagreement about what constitutes positive change, we have included an academic-behavioral outcome called "career and life goal development." The implication is that the student should gain in ability to explore, consider, and make decisions about eventual goals.

### Form of Outcome Measures

For many institutions, there may be an inherent conflict between observing Bowen's caveat about feasibility of measurement and adopting

Astin's value-added approach, which statistically controls for student entry characteristics when observing changes in student outcomes over time. This is particularly true if measures of cognitive development, such as reasoning skills and critical thinking, are used to supplement more traditional measures of academic achievement. In developing new measures and in assisting institutions with the use of already developed measures, NCRIPTAL will attempt to help simplify the appropriate use of outcome measures.

The next section of this paper describes some of the academic measures already in use by colleges and alerts the reader to some new measures that NCRIPTAL staff hope to make available for future use.

# III. Outcome Measures and Outcome Research

As discussed in the previous sections, out come measurement recently has received increased emphasis at the college level. Numerous measures are available to assess learning in college. It is difficult for educators to choose among the widely diverse types of measures. In this section, some of the available measures will be reviewed. Reliability and validity information are included when available and scholarly research that has been conducted using the measures is reported. The purpose is to describe the utility of the instrument for measuring improved learning and teaching by examining the measure's properties and the results it has produced as a college outcome measure. New measures and new reports on their uses are appearing daily. This review should be considered background for future updates.

The measures will be divided into three sections consistent with the cells in NCRIPTAL's typology: academic-cognitive outcomes, academic-motivational outcomes, and the academic-behavioral outcomes.

**Academic-Cognitive Outcomes**

The following available measures are reviewed here:

Graduate Record Examination
American College Testing Program Achievement Tests
Undergraduate Achievement Program of the College Entrance Examination Board
The American College Testing Program College Outcome Measurement Program (COMP)
College Level Examination Program
National Teacher Examination
Measures of critical thinking
Measures of basic skills

*Graduate Record Exam*

The Graduate Record Exam, produced by the Educational Testing Service, was initially introduced as a general achievement test to measure knowledge in three general categories: social studies, natural sciences, and humanities. These tests were different from the typical achievement tests of that era because the items were meant to evaluate students' ability to read, understand, and interpret knowledge rather than to test simply their possession of knowledge. Though the items were somewhat content-imbedded, the initial area tests of the GREs were developed to test ability to generalize from information that was given (Pace, 1979).

The general test of the GREs is a test of developed verbal, quantitative, and analytical abilities that have been acquired by students over time. The GRE general test is offered to college seniors and graduates and is used by some graduate schools for admission decisions, fellowship awards, and prediction of an applicant's success in graduate school.

Educational Testing Service also produces twenty GRE advanced subject tests that measure knowledge specific to certain fields. These tests are intended for college seniors and are fairly comprehensive. These scores also are used for admission criteria in some graduate schools.

The K-R 20 reliability coefficients for verbal and quantitative exceed .90 and for the analytical section are .86 (Cohn, 1985). Though these values are highly respectable, Jaegar (1985) warns that an internal consistency measure such as the K-R 20 may actually overestimate the reliability of the general sections.

The validity of the GRE general tests is somewhat questionable since there is little evidence to support the predictive power of the tests for graduate school achievement. The validity coefficients for predicting first-year grade-point averages for the three sections of the general tests are around .20 and .30 (Cohn, 1985). It should be noted, however, that though these correlations are only moderate, the sample of students is limited to those who have been accepted into a graduate program. Thus both the range of scores and the number of students included in the sample are small.

In combination with undergraduate grade point average, the predictive validities of the GRE general test for graduate school success range from .32 to .56 (Jaegar, 1985). Because ETS encourages the use of GRE scores in conjunction with other admission criteria (e.g., G.P.A., letters of recommendation), this combined validity justifies its use.

Numerous researchers have used the GREs as a measure of differences in teaching and learning at both the institutional and individual levels. Many of these studies have been reviewed by Pascarella (1985) and we have drawn freely from that review.

At the individual level, Nichols (1964) attempted to assess the effects of different colleges on the GRE verbal and quantitative scores of

students. Nichols examined structural and organizational characteristics of colleges (private vs. public, faculty-student ratio, enrollment, and library books per student) and the environmental characteristics (using Astin's 1963 Environmental Assessment Technique). The college structural characteristics were not significantly correlated with the GRE scores but three aspects of the EAT were significantly correlated with the verbal and quantitative scores. The amount of variance accounted for by these EAT variables, however, was small in comparison to the amount of variance accounted for by the students' entry characteristics.

Astin (1968) examined variation on the humanities, natural science, and social science GRE tests as a function of traditional indices of institutional quality. These indices included intelligence of student body, financial resources, library size, and student-faculty ratio. Astin found, however, that all of the partial correlations indicating a relationship between institutional characteristics and GRE scores became trivial after accounting for numerous student entry characteristics (e.g., aspirations, high school achievement, family background). Apparently, student characteristics are more predictive of GRE area scores than institutional characteristics. This finding indicates that changes in learning may not be attributed to institutional characteristics, but perhaps must be examined at a lower programmatic level.

In further analyses of the same data, Astin and Panos (1969) found that institutional characteristics other than traditional quality indices explained some GRE variance. These characteristics included institutions where students made frequent use of automobiles and where students were undecided about their careers. Also, GRE scores were higher at institutions where there was a generally flexible curriculum, where there was a technical emphasis, and where there was a large enrollment. These partial correlations, however, were also quite small, indicating that pre-enrollment characteristics may be more meaningful than institutional characteristics for determining college outcomes.

At the institutional level, Rock, Centra, and Linn (1970) and Centra and Rock (1971) attempted to explain the relationship between college characteristics and student learning. Their dependent measure was residual scores on the three area tests of the GREs, the humanities, social sciences, and natural sciences. To obtain these residuals, the authors regressed the average institutional GRE score on the average SAT score which yielded predicted GRE scores for each institution. The predicted scores were then subtracted from the actual scores which produced the residual score. The proportion of students majoring in the various areas was also taken into account.

Rock, Centra, and Linn (1970) examined the influence of institutional characteristics typically associated with quality on GRE residual scores. Only two of the factors, the income a college receives per student and the proportion of faculty with a doctorate, were consistently related to colleges with high residual achievement.

Centra and Rock (1971) focused on the differences between environmental characteristics of colleges and their potential influence on learning. The five factors they used, which were derived from the Questionnaire on Student and College Characteristics, were faculty-student interaction, curriculum flexibility, cultural facilities, student activism, and degree of academic challenge. Centra and Rock found a positive relationship between achievement and faculty-student interaction, curricular flexibility and availability of cultural activities.

In his review of large-scale unpublished surveys, Pace (1979) discusses two studies that examine academic achievement during college using the GREs as the outcome measure. An ETS compilation of 3,035 scores of seniors from various colleges showed that students who majored in one of three subareas (social science, natural science, and humanities) scored higher on that section the the GRE area tests than students who majored in another area. Pace states that the results simply attest to the fact that "students know most what they study most" (p. 25).

The second study Pace reviewed involved the advanced tests of the GREs. Harvey and Lannholm (1960) tested 300 upperclassmen at 29 institutions both before the students had taken any upper level division courses and again at the end of their senior year. Students who had majored in psychology, economics, or chemistry were included in the sample. The differences in scores were typically close to a standard deviation higher after having taken the upper level courses. This evidence supports the contention that students learn from studying in a specific field.

These studies seem to reveal that students learn in college and the more they study in a certain field, the more they learn in that field. The GRE tests appear to be a reasonable measure for examining learning at the college level. While few studies have conducted pre- and post-tests of college students' learning using the GREs, this option appears to have potential for a useful measure of differential teaching and learning.

*American College Testing Program*

The ACT tests are designed to measure educational development in the areas of mathematics usage, English usage, social studies reading, and natural sciences reading. The ACT Assessment Program also includes an Interest Inventory and a Student Profile. The program was developed in 1954 as a college admissions test and as a tool for guidance and counseling of freshman in college. The program originally grew out of the Iowa Tests of Educational Development but has since then become independent.

Five educational development scores are reported after the ACT tests have been taken. Four individual scores are reported for the four subsections and a composite score representing an average of the four sections is also included.

The content validity of the tests is acceptable and reasonable according to two reviewers (Aiken, 1985; Kifer, 1985). Predictive validity for the ACT tests is also quite high. The validity ranges from .4 to .5 with college freshman grade- point average (Aiken, 1985). However, when high school grade point average is already included in the regression equation, the inclusion of ACT scores improves the predictive validity by only .10. Kifer (1985) questions the value of the effort invested to obtain this limited increase in predictive power.

Another issue mentioned by numerous reviewers is the amount of overlap within the four sections of the ACT (Hill, 1978; Kifer, 1985; Aiken, 1985). There is agreement among these critics that too much emphasis is placed on reading in the various sections of the test. The intercorrelations of the four sections range from .53 to .68, indicating that similar abilities are being tested.

The reported reliabilities for the various subsections are as follows: English usage, .92; mathematics usage, .91; social studies reading, .88; and natural sciences reading, .88 (Aiken, 1985). These reliabilities have improved over the years and are adequate at these levels for individual decisions based on test scores.

The ACT test scores have been used by researchers to examine learning and cognitive development at the college level. Lenning, Munday, and Maxey (1969) examined cognitive growth in the first two years of college in five institutions using tests of the American College Testing Program. Samples of students were chosen from two state colleges, one liberal arts college, a junior college, and a state university. ACT tests were administered at the beginning of the freshman year and again at the end of the sophomore year. Differences between pre- and post-test scores were significant on all the composite scores for all groups except the female sample at one institution. Students made the greatest gains in social studies and natural sciences and somewhat lesser gains in English and mathematics.

Dumont and Troelstrup (1981) also used the ACT tests as a measure of cognitive gains in college. They pre-tested students at one institution at the beginning of their freshman year and again four years later. Students made significant gains in all areas, showing even more of an increase in ACT subscores than the sophomores in Lenning, Munday, and Maxey's sample.

The use of pre- and post-tests for students highlights the actual learning that occurs in college. The ACT tests seem capable of tapping the cognitive development of students in general education areas.

*Undergraduate Assessment Program: Area Tests and Field Tests*

The Undergraduate Assessment Program (UAP) is closely related to the GRE. The UAP area tests are fundamentally the same as the GRE area tests and are similar to the GRE advanced tests, however, the Business field test is the only field test still available (Pace, 1979). The area tests are divided into three sections: humanities, social sciences, and natural sciences.

At part of the standardization of the UAP test, 47,000 seniors from 211 colleges were given the area tests. Some of the colleges also administered the tests to other classes. Despite the problems of comparing cross-sectional data, especially when the population of colleges was so heterogeneous and the sample sizes so different, the results showed that, within the three major domains, seniors and juniors scored higher than sophomores and freshmen (Pace, 1979).

Further results from ETS (1976) indicate that when the UAP test is in the student's area of interest (i.e., humanities, social sciences, or natural sciences), the scores are substantially higher than the scores of the total group of students from that institution (including the 'interest' group in the total). Seniors within their area of interest scored higher than sophomores having the same academic interest (ETS, 1976).

ETS revised the area tests in 1978 and published a new guide (ETS, 1978). This new guide reports results similar to the previous one. Seniors who majored in humanities, social science, or natural science scored considerably higher than sophomores and freshman. The mean scores on the three area tests increase with each year of college as well (remember that this is a cross-sectional data set). Thus, the evidence reported by ETS leads one to believe that the more one concentrates on a particular field, the

more one learns in that field. Students are learning in college and the UAP tests seem able to capture some aspects of students' cognitive development.

*ACT College Outcome Measures Project*

The American College Testing Program has developed a unique achievement test called the College Outcome Measures Program. This project represents a new direction in achievement testing (Pace, 1979). Rather than testing general knowledge and specific content from an academic discipline, the COMP is an effort to measure students' ability to apply facts, concepts, and skills to real world activities. Specifically, the content of the Measurement Battery involves three areas related to adult functioning: functioning within social institutions, using science and technology, and using the arts. Three types of competencies are measured within these three areas: communicating, solving social problems, and clarifying social values.

The COMP is unusual in its format and in the materials used for testing. Unlike the typical multiple-choice, paper and pencil format, the COMP materials include film excerpts, taped newscasts, art prints, magazine and newspaper articles, and other realistic materials that might be encountered in life. An actual item on the COMP might require a recorded or a written response: the student may have to write a persuasive memo and justify a decision in speech. The idea is to make the test as realistic and as relevant to real life as possible.

The COMP, therefore, takes considerable time to administer as well as to evaluate and score. Taking the test requires approximately six hours and rating the responses takes about one hour. Standardized rating scales have been developed to aid judges in their evaluation process (Forrest & Steele, 1978).

In addition to the six-hour battery, the overall COMP includes two additional tests: an Activity Inventory and an Objective Test. The Objective Test is an effort to maintain the advantages of the Measurement Battery while decreasing the time factor. This test uses the same stimuli as the Measurement Battery but the student is given four options from which to choose, two of which are considered good answers. This test can be machine scored and test-taking time is reduced to two and one-half hours.

The Activity Inventory measures the amount of experience that an individual has had in the six areas that are covered in the Measurement Battery. The score is meant to supplement either the Objective Test or the Measurement Battery with an experience factor.

*College Level Examination Program*

The College Level Examination Program (CLEP) is a program originally developed for giving college credit to students by examination. CLEP now offers two types of examinations: the General Examinations and the Subject Examinations. The examinations were developed to assess the knowledge of students who have acquired knowledge outside the classroom. The General Exams measure college-level achievement in English composition, humanities, mathematics, natural sciences, and social sciences and history. The test is for general education requirements and covers material typically studied in the first two years of college. The Subject Examinations are more advanced and require specific knowledge in a particular field. Only the General Exams will be discussed here.

Reliability coefficients for the General Exams are quite high, on the order of .90 and above (Aleamoni, 1985). Validity of the exams, however, is somewhat in question. The primary validity information comes from data used for norming the tests. These data showed that the more courses taken in an area (i.e., humanities, history and social sciences, science, and math), the higher the score on the test (Pace, 1979). These validity tests were based on a national sample of college sophomores (N = 2600). The manual from which information was made available, however, did not include the correlations between number of courses taken and test scores. The fact that a relationship exists is fairly weak evidence for validity. Aleamoni (1985) states that many colleges have had to develop their own validation studies in order to make a case for the appropriateness of the exams.

*National Teacher Examination*

The National Teacher Examination initially produced by the Educational Testing Service was a test designed for college seniors and teachers and provided a standardized measure of academic preparation in three areas: general education, professional education (for teachers), and subject-field preparation. The purpose of the test was threefold: it could be used to assist colleges in reviewing their programs and policies; state departments could use the scores for teacher certification purposes as well as for attaining profiles of prospective teachers' knowledge and skills; and school administrators could use the scores as a standardized measure for evaluating the competencies of prospective teachers. ETS, however, warned against the use of these exams as a sole determinant of graduation, certification, and selection decisions. Colleges are warned in

the exam booklet against using absolute cut-off scores for any type of decision (Merwin, 1978).

As described above, the NTE actually is two exams: the Common Examination and the Area Examinations. Only the Common Examination will be discussed here. The Common Examination covers general education and professional education: the professional education section consists of 110 items, the general education sections consist of 45 items on Written English Expression, 65 items on Social Studies, Literature, and the Fine Arts, and 50 items on Science and Mathematics. These sections are generally seen as including knowledge a well-rounded educator should possess.

Because of the increasing concern with teacher certification and the emerging needs of institutions, ETS expanded the National Teacher Examinations just described into the National Teacher Examination Program. This new program is composed of three sections: the Pre-Professional Skills Tests (PPST), the NTE Core Battery, and the NTE Specialty Area Tests.

The PPST is an assessment of basic skills initially developed for use in colleges to determine whether a student had the basic skills necessary to enter a teacher training program. There has been an increase, however, in the use of the PPST as an initial teacher certification test. The PPST measures skills required for the beginning teacher. The battery consists of three tests: Communications Skills, General Knowledge, and Professional Knowledge. The Communications Skills test covers listening, reading, and writing abilities. The General Knowledge test covers mathematics, science, social studies, literature, and fine arts. The Professional Knowledge test covers knowledge and skills needed for developing instructional plans and their implementation as well as the professional behavior required of teachers. The NTE Specialty Area tests are content specific tests available in 28 areas.

Recent research by Ayres and Bennett (1983) and Ayres (1983) used the old NTE as an outcome variable to assess differences in learning across institutions and individuals. The authors judged the NTE to be a reasonable measure of learning usually expected during general undergraduate study. Using the institution as a unit of analysis (N = 15) Ayres and Bennett (1983) explained 88% of the variance in NTE scores by including in the regression equation average institutional SAT score, average number of courses taken in general education, average faculty salary, average educational attainment of faculty, institutional age, library size, and institutional size. The average educational attainment of faculty members accounted for the largest percentage of the variance.

Using the same data, but examining it with the student as the unit of analysis (N = 2,229), Ayres (1983) investigated the effects of the racial composition of the institution on NTE scores. Ayres found that when controlling for aptitude (SAT score), black students in a primarily white institution performed better than black students in a predominantly black institution.

These successful attempts at explaining variation on the NTE provide evidence for its usefulness as a measure of some aspects of college achievement. The evidence presented by ETS (in Merwin, 1978) suggest, on the other hand, that the predictive validity of the NTE limits its usefulness as a selection or certification tool in education. As an academic outcome measure of the effectiveness of teaching and learning, however, it may be useful and appropriate.

### *Critical Thinking and Higher Level Outcome Measures*

A number of tests have been developed to measure the higher level cognitive processes of college students. These cognitive processes include critical thinking, complex reasoning and judgment, abstract thinking, and flexibility of thought. In this section, some research findings are briefly reviewed and information available on the tests used are reported. The measures used in this research, however, are often not standardized and are developed by researchers for their particular interests. Neither can we claim to have exhausted the extensive and growing literature in this area. Within the NCRIPTAL work, McKeachie and colleagues present a more complete review of such measures.

Evidence supports the notion that critical thinking abilities improve over the college years. Lehmann (1963) used the American Council of Education's Test of Critical Thinking in a longitudinal study of students at Michigan State University. The ACE test taps five dimensions of critical thinking: (1) defining a problem, (2) selecting information relevant to the problem, 93) recognizing stated and unstated assumptions, (4) formulating and selecting relevant hypotheses, and (5) drawing valid conclusions. Lehmann tested 1,051 students on entering college and again at the end of the freshman year and every subsequent year. All students tested had significantly higher scores as seniors than they did as freshman. The most significant gains occurred during the freshman year.

Keeley, Brown, and Kreutzer (1982) used a cross-sectional design and administered open-ended and essay measures of critical thinking to 145 freshman and 155 seniors at a large state univer-

sity. Two experimental conditions were employed, with half of each class cohort receiving each treatment. In one condition, students were given very general instructions on how to respond to the items. The other group received specific instructions for writing critical evaluations of the items.

In the general instruction condition, seniors had significantly higher scores in six of the seven criticism categories on which they were judged. These categories were general criticisms, understanding of structure, logical inconsistency, explicit criticism, and essay length. In the specific instruction condition, seniors were more skilled at identifying the controversy and conclusions of the essay, and identifying assumptions. The seniors also received higher overall scores in this condition.

Another critical thinking measure, the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964), has been used by a number of researchers and may be the most widely used critical thinking instrument. This instrument is designed to measure three dimensions of critical thinking: inference, deduction, and recognition. Recognition refers to the ability to recognize unstated assumptions. Inference refers to the ability to distinguish between valid and invalid inferences drawn from data. Deduction refers to the ability to reason deductively, from the general to the specific.

Mentkowski and Strait (1983) studied the development of critical thinking skills during the college years using the Watson-Glaser. This research was part of a comprehensive outcome evaluation program that Alverno College uses to examine cognitive development of students.

The design was longitudinal, with more than 700 freshman tested at the beginning of their freshman year and again at the end of the sophomore and senior years. Significant increases in scores on all three dimensions of the Watson-Glaser were found between sophomore and senior years. Significant increases on the inference and deduction scales were found between the freshman and sophomore years.

An additional assessment tool used in Alverno's Evaluation Program is a Piagetian formal reasoning task. The task essentially measures the student's ability to reason abstractly: there were two proportionality problems, two conservation of volume problems and one problem dealing with the separation of variables.

Mentkowski and Strait (1983) found significant increases in formal reasoning between freshman and sophomore year. A similar increase in formal reasoning ability was also found by Eisert and Tomlinson-Keasey (1978) in a study of 55 freshman. They found significant increases between the start and end of the freshman year.

Another measure available is the Test of Thematic Analysis (Winter & McClelland, 1978). This broad essay exam measures thinking and reasoning ability. In the test, students are given two different groups of Thematic Apperception Test stories and are asked to describe the differences between the two groups in an essay. The essays are judged on nine reasoning and thinking criteria.

Winter and McClelland (1978) conducted a multiple institution study of cognitive development using the Test of Thematic Analysis. Samples were drawn from three institutions: an elite liberal arts college, a state teachers college, and a community college. Longitudinal and cross-sectional data were collected from the liberal arts college; the design for the teachers college and community college was cross-sectional.

Longitudinal data from 80 students at the liberal arts college showed significant differences from freshman to senior year on the Thematic Analysis score. The cross-sectional designs resulted in significant findings only at the elite liberal arts college. Data from the teachers college and community college did not show significantly reliable increases.

Winter, McClelland, and Stewart (1981) sought to measure intellectual flexibility in reasoning using Stewart's Analysis of Argument Test (1977). This test confronts a subject with a controversial statement and asks the subject to write two essays: one defending the statement and one attacking it. The two essays are scored on ten criteria. These criteria are meant to measure the extent to which the subject can evaluate an argument and construct a coherent evaluation of an argument.

The study involved the same samples as those in Winter and McClelland's 1978 study. Statistically significant differences were found between freshman and final-year students at all three institutions on total score of the Analysis of Argument Test.

The reflective judgment interview (RJI) is another available measure of higher level cognitive skills. Reflective judgment refers to the development of complex reasoning and judgment skills. In the interview, the subject is confronted with four controversial dilemmas and a set of standardized questions designed to tap level of reasoning. Level of reasoning is determined based on a Perry-like scheme of intellectual development (1970). Schmidt and Davison (1981) classify level of reasoning along a multilevel continuum, ranging from dualism to probabilism. Dualism is a simple and illogical reasoning pat-

tern and probabilism is reasoning based on evidence and logic.

Brabeck (1983) provides a review of ten studies that have used the RJI as a dependent variable. Many of these studies used a cross-sectional design to measure differences in reflective judgment across educational levels. The results support the idea of a reflective judgment continuum, with more advanced students showing higher levels of reasoning ability. Postsecondary education does have an influence on the development of reasoning ability.

NCRIPTAL researchers, McKeachie, Pintrich, Lin, and Smith provide a detailed and more technical review of measures of two other academic-cognitive outcomes, problem- solving skills, and knowledge representation. Their investigation will be incorporated into this background paper in its next revision. Readers are referred to the list of NCRIPTAL technical reports noted at the beginning of this report.

In sum, many higher level cognitive processes appear to improve in college. Many measures are available for evaluating improvement in these areas. In this area, most measures are developed specifically for a given research program though they may be useful for other purposes.

*Basic Skills*

Although basic skills are generally not viewed as an outcome of college, they are becoming an increasingly important part of the college curriculum. As access to college for disadvantaged students has improved, so has the concern for keeping those students in college and helping them to succeed through the use of remedial and basic skill programs.

Concern for basic skills programs at the four-year college level has grown out of fairly recent changes in the student population. Two-year community and junior colleges have always been concerned with basic skills as both a requirement and an outcome of their programs.

For the purposes of evaluating outcomes as well as the maintenance of skills in both two- and four-year colleges, basic skills measures are considered here as an outcome measure that some colleges may wish to incorporate or use as a pre-test at entry.

Basic skills usually refer to fundamental abilities in mathematics, English composition, reading, and vocabulary. While most programs for identifying basic skills at the college level are developed at the institutions, there are tests available for evaluating students in this area. Appendix B presents information on a number of reading, math, and vocabulary tests designed for the college level. While not an inclusive list, it represents some available tests of basic skills.

Wolfe (1983) studied the development of basic skills in college. He investigated the progression of students in vocabulary and mathematical ability using 1979 follow-up data from the National Longitudinal Study of the High School Class of 1972. He found that, when controlling for ethnicity, parent's education, father's occupation and the 1972 scores, postsecondary education significantly improved both vocabulary and mathematics performance.

**Academic-Motivational Outcomes**

Measures covered in this section include those related to: motivation, self-efficacy, student involvement, and satisfaction with college.

Two motivational outcomes, motivation and self-efficacy, are discussed in detail in a companion report by McKeachie et al. Another companion paper by Korn discusses self-efficacy as both an input and outcome variable in college. Rather than repeat their efforts, we refer the reader interested in these outcomes to the two other reviews.

*Involvement and Quality of Effort*

The relatively new concepts of involvement and quality of effort have both motivational and behavioral components. We have chosen to include them as motivational outcomes because of the ties between these concepts and motivation.

Astin's (1984) theory of student involvement can be simply summarized as "Students learn by becoming involved" (Astin, 1985, p. 133). Astin defines involvement as the amount of physical and psychological energy that the student devotes to the academic experience. He considers involvement as closely linked to motivation but prefers the term involvement because it has more behavioral implications.

Involvement theory has five postulates. First, involvement consists of investing energy into objects. Second, involvement exists along a continuum. Third, involvement has quantitative and qualitative aspects. Fourth, the amount of student learning and personal development is related to the amount of student involvement. Fifth and finally, the effectiveness of an institutional policy or practice is related to the capacity it has to increase student involvement.

Astin (1977) conducted a large-scale longitudinal study investigating the effects of various forms of involvement on numerous student outcomes. His general conclusion was that most forms of involvement lead to greater than average

changes in the entry characteristics of freshmen. In some instances, involvement was more strongly related to outcomes than either entry or institutional characteristics.

Most of the outcomes Astin measured were motivational outcomes. More research is needed to determine the effects of involvement on cognitive outcomes. No obvious instrument is yet available to measure student involvement.

Pace (1984) discusses the concept of quality of effort as an important determinant of student outcomes. He posits that because education is both a process and product, the quality of the educational experience must be taken into account. The quality of this process is not the sole responsibility of the institution or its faculty members, rather students must take some responsibility for their own progress by taking advantage of opportunities provided to them by the institution. Thus, by measuring quality of student effort one can better assess the quality of the educational process.

Pace developed an instrument to measure the quality of student experiences by determining the extent to which students take part in activities and opportunities intended to promote student learning and development. Pace's Quality of College Student Experiences (1984) standardized self-report survey includes fourteen scales of activities (e.g., student union, athletic and recreational facilities, experiences in writing, library experiences) which reflect increasing amounts of effort and potential value. The scored responses provide a measure of the quality of effort students have invested in the various aspects of college life.

The survey also collects information on college environment, student background information, and gains made during college. Pace suggests that the instrument be used by institutions for program evaluation, resource allocation, and faculty and staff discussions. The instrument can also be used as a research tool for investigating the relationship between quality of student effort and institutional characteristics.

Pace (1984) reports reliability and validity information in the manual. For reviews of the instrument and comments on the psychometric features reported in the manual see Miller (1985) and Brown (1985).

Pace also reports data from studies using the Quality of College Student Experiences Survey. Results support the idea that quality of effort is an important predictor of student achievement; effort measures significantly increase the amount of explained variance in student achievement (Pace, 1984). The increase in explanatory power occurs when student characteristics, college status variables, and college environment ratings have already been included in the regression equation.

The concepts of quality of effort and involvement are often considered process variables; that is, variables that moderate the relationship between available learning opportunities and student outcomes. However, when education is viewed as an iterative process with current outcomes influencing future achievement, then these concepts can be considered outcomes in the sense that the development of effort and involvement will be an outcome that in turn influences the educational process. For this reason, we are including these as possible outcome measures at the college level.

**Academic-Behavioral Outcomes**

Behavioral measures of students' cognitive development are difficult to find in the literature. Most measures used in research and practice are either motivational or cognitive. As mentioned earlier, involvement and quality of effort measures can be considered behavioral in the sense that they measure self-reports of participation in various activities. Other behavioral measures of academic outcomes that NCRIPTAL researchers are considering include goal-exploring behaviors, relationships with faculty, and persistence.

*Career and Life Goal Exploration*

One desired outcome of college is the exploration and development of life options, including both appropriate career choices and recognition of values to be gained from liberal education. Interestingly, however, literature in these areas seems to advocate one of these types of student development to the exclusion of the other.

A wide variety of instruments exists to measure students' career exploration activities. These include vocational development inventories, career maturity scales, and search procedures that help students identify occupational groups with similar personality characteristics or interests. Most of these instruments are designed to assist the "undecided college student" and appear to be based on the assumption that being undecided is both economically inefficient and psychologically unsettling to the college student.

In a different mode, considerable rhetoric advances the value of liberal education in preference to early (or premature) decisions about career specialization. Although this openness to liberal learning is highly valued by many educators, the authors know of no instrument designed to measure such student proclivities. Most existing information is based on surveys of entering college students. In recent years the percentage of students espousing vocational goals has risen substantially while the percent who desire general education remains relatively stable.

From the standpoint of improving teaching and learning, the important question is whether student openness to considering various educational and career alternatives changes because of specific educational experiences. For example, does study of liberal arts subjects result in students' placing greater or less value on this knowledge? Does studying career-oriented subjects close one's mind to the value of liberal education?

Such questions have been studied by only a few researchers. To illustrate, Mentkowski and Doherty (1984) report that students at a college with a competency-based liberal arts curriculum moved from strong initial career orientations toward an appreciation of liberal arts.

Instruments that measure student goals in a multi-dimensional fashion so that change in either direction can be assessed await development.

*Exploration of Diversity*

We have used the term "exploration of diversity" to represent a complex set of educational outcomes that are not captured in other categories of our framework. One of the best known frameworks for measuring such outcomes in the academic-cognitive sense is Perry's scheme of intellectual development (Perry, 1970). In this scheme, intellectual development is measured by students' movement from a position of dualism (right/wrong) to a more balanced consideration of a variety of viewpoints (relativism), finally to selecting and justifying one's own point of view (commitment). A variety of paper and pencil measures of student change on the Perry dimensions are under development.

In the academic-behavioral sense, exploration of diversity may be reasonably well captured by some portions of Pace's Quality of Student Experience Scale, discussed earlier. Attendance at campus events, for example, might be a measure of the impact of college in broadening student horizons as well as an index of student effort.

Our intent in listing these exploratory behaviors as outcomes is to stimulate thinking about broad behavioral observations through which colleges might measure the extent to which students become more likely to participate in further education and cultural affairs and to exhibit other behaviors generally attributed to educated people. One of the deficiencies in previous outcome typologies has been strong dependence on high inference measures and a notable lack of actual behavior observations in assessing student outcomes. We will be exploring measures of these types of outcomes in the future.

*Persistence*

Why students drop out of college has been studied by numerous researchers in an effort to understand the determinants of attrition. Tinto's (1975) model of attrition posits that academic and social integration of the student into the institution and the students' interaction with these systems are the primary determinants of persistence in college.

In testing the Tinto model, Munro (1981) found that academic integration had a strong effect on persistence while social integration was not a significant predictor. Similarly, Pascarella, Duby, and Iverson (1983) also found that academic integration was a significant predictor of student persistence. In addition, they found that entry characteristics were more predictive of persistence in a non-residential setting than in a residential setting.

Pascarella, Smart, and Ethington (1986) studied persistence in students at two-year colleges over a period of nine years. They found that both academic and social integration were important predictors of persistence when the students were tracked for a longer period of time.

Edwards and Waters (1983) attempted to explain persistence by using academic course involvement, academic ability, academic performance, and satisfaction with both courses and college in general as predictors. In a replication study, when they included a personal needs/college climate discrepancy index and a voluntary/involuntary attrition breakdown, they found that the discrepancy index was marginally significant as a predictor of voluntary attrition.

As an outcome variable, persistence thus has most often been correlated with various independent variables in hopes of identifying facilitating conditions. While it is not a direct measure of improved learning at the college level, attendance is a prerequisite for continued cognitive development that can be directly linked to the college experience. Quite possibly, persistence as a dichotomous variable is not as useful an outcome in achieving teaching and learning improvement as involvement or quality of effort, which could be construed to represent various levels of persistence.

*Faculty-Student Relationships*

An additional behavioral outcome to be considered here, student interactions with faculty, can be considered both an outcome and process variable. Informal interactions with faculty are included in Tinto's model of attrition as an important aspect of academic integration. In this sense, relationships are considered process

variables in the educational cycle; they determine and affect educational outcomes. However, when viewing education as an iterative process, current relationships with faculty can be viewed as outcomes that may affect future educational outcomes.

Pascarella (1980) completed a comprehensive review on the relationship between informal student-faculty interaction and college outcomes. He concluded that the extent and quality of student-faculty interactions had significant positive associations with students' educational aspirations, their attitudes toward college, their academic achievement, their intellectual and personal development, and their persistence in college. Bean and Kuh (1984), on the other hand, found no significant relationship between informal contact with faculty and student grade point average.

A recent study by Volkwein, King, and Terenzini (1986) investigated the relationships that develop between faculty members and transfer students. In this subsample of college students, perceptions about the quality and strength of their relationships with faculty were significantly related to intellectual growth.

Thus far, most measures of faculty-student relationships have been student self-reports of the number of hours per semester of non-class-related interactions with faculty. There is room for development of other valid measures of this association.

In sum, there are several academic-behavioral outcomes of college but few measures have been developed. Though these outcomes have been investigated less frequently than the cognitive and affective outcomes, there is evidence to support the importance and salience of these outcomes for college students.

**Conclusion**

In this working paper, NCRIPTAL has specified its concerns for the technical parameters of outcome measures that are part of the current discussions of assessment, assuming that the policy parameters are held constant through our collaborative work with institutions that desire to improve teaching and learning. Further, we have delineated the type and form of outcomes through which we hope to measure the effectiveness of various alterations in teaching and learning environments. We have briefly reviewed some of the forms of outcome measurement that are available for our use and that of others, and we have identified some gaps in available measurement techniques. As we learn more about new instruments and techniques that are being used with apparent success in purposive improvement of teaching and learning, we will expand the information in this paper.

# Appendixes

## Appendix A. Categories of the NCHEMS Outcome Structures

**Categories of the NCHEMS Outcomes Structure**

| CAT. CODE • | ENTITY BEING MAINTAINED OR CHANGED |
|---|---|
| **1000** | **Economic Outcomes** |
| 1100 | Economic Access and Independence Outcomes |
| 1110 | Economic Access |
| 1120 | Economic Flexibility, Adaptability, and Security |
| 1130 | Income and Standard of Living |
| 1200 | Economic Resources and Costs |
| 1210 | Economic Costs and Efficiency |
| 1220 | Economic Resources (including employees) |
| 1300 | Economic Production |
| 1310 | Economic Productivity and Production |
| 1320 | Economic Services Provided |
| 1400 | Other Economic Outcomes |
| **2000** | **Human Characteristics Outcomes** |
| 2100 | Aspirations |
| 2110 | Desires, Aims, and Goals |
| 2120 | Dislikes, Likes, and Interests |
| 2130 | Motivation or Drive Level |
| 2140 | Other Aspirational Outcomes |
| 2200 | Competence and Skills |
| 2210 | Academic Skills |
| 2220 | Citizenship and Family Membership Skills |
| 2230 | Creativity Skills |
| 2240 | Expression and Communication Skills |
| 2250 | Intellectual Skills |
| 2260 | Interpersonal, Leadership, and Organizational Skills |
| 2270 | Occupational and Employability Skills |
| 2280 | Physical and Motor Skills |
| 2290 | Other Skill Outcomes |

NOTE. The fourth-level categories, into which any of the categories listed here can be divided, are "maintenance" (a fourth digit of "1") and "change" (a fourth digit of "2").

SOURCE. Oscar T. Lenning, Young S. Lee, Sidney S. Micek, and Allan L. Service, *A Structure for the Outcomes of Postsecondary Education* (Boulder, Colo.: National Center for Higher Education Management Systems, 1977), p. 27.

Appendix A *(continued)*

**Categories of the NCHEMS Outcomes Structure,** *continued*

| CAT. CODE # | ENTITY BEING MAINTAINED OR CHANGED |
|---|---|
| **2000** | **Human Characteristics Outcomes,** *continued* |
| 2300 | Morale, Satisfaction, and Affective Characteristics |
| 2310 | Attitudes and Values |
| 2320 | Beliefs, Commitments, and Philosophy of Life |
| 2330 | Feelings and Emotions |
| 2340 | Mores, Customs, and Standards of Conduct |
| 2350 | Other Affective Outcomes |
| 2400 | Perceptual Characteristics |
| 2410 | Perceptual Awareness and Sensitivity |
| 2420 | Perception of Self |
| 2430 | Perception of Others |
| 2440 | Perception of Things |
| 2450 | Other Perceptual Outcomes |
| 2500 | Personality and Personal Coping Characteristics |
| 2510 | Adventurousness and Initiative |
| 2520 | Autonomy and Independence |
| 2530 | Dependability and Responsibility |
| 2540 | Dogmatic/Open-Minded, Authoritarian/Democratic |
| 2550 | Flexibility and Adaptability |
| 2560 | Habits |
| 2570 | Psychological Functioning |
| 2580 | Tolerance and Persistence |
| 2590 | Other Personality and Personal Coping Outcomes |
| 2600 | Physical and Physiological Characteristics |
| 2610 | Physical Fitness and Traits |
| 2620 | Physiological Health |
| 2630 | Other Physical or Physiological Outcomes |
| 2700 | Status, Recognition, and Certification |
| 2710 | Completion or Achievement Award |
| 2720 | Credit Recognition |
| 2730 | Image, Reputation, or Status |
| 2740 | Licensing and Certification |
| 2750 | Obtaining a Job or Admission to a Follow-up Program |
| 2760 | Power and/or Authority |

Appendix A *(continued)*

**Categories of the NCHEMS Outcomes Structure,** *continued*

| CAT. CODE # | ENTITY BEING MAINTAINED OR CHANGED |
|---|---|
| **2000** | **Human Characteristics Outcomes,** *continued* |
| 2770 | Job, School, or Life Success |
| 2780 | Other Status, Recognition, and Certification Outcomes |
| 2800 | Social Activities and Roles |
| 2810 | Adjustment to Retirement |
| 2820 | Affiliations |
| 2830 | Avocational and Social Activities and Roles |
| 2840 | Career and Vocational Activities and Roles |
| 2850 | Citizenship Activities and Roles |
| 2860 | Family Activities and Roles |
| 2870 | Friendships and Relationships |
| 2880 | Other Activity and Role Outcomes |
| 2900 | Other Human Characteristic Outcomes |
| **3000** | **Knowledge, Technology, and Art Form Outcomes** |
| 3100 | General Knowledge and Understanding |
| 3110 | Knowledge and Understanding of General Facts and Terminology |
| 3120 | Knowledge and Understanding of General Processes |
| 3130 | Knowledge and Understanding of General Theory |
| 3140 | Other General Knowledge and Understanding |
| 3200 | Specialized Knowledge and Understanding |
| 3210 | Knowledge and Understanding of Specialized Facts and Terminology |
| 3220 | Knowledge and Understanding of Specialized Processes |
| 3230 | Knowledge and Understanding of Specialized Theory |
| 3240 | Other Specialized Knowledge and Understanding |
| 3300 | Research and Scholarship |
| 3310 | Research and Scholarship Knowledge and Understanding |
| 3320 | Research and Scholarship Products |

Appendix A *(continued)*

**Categories of the NCHEMS Outcomes Structure,** *continued*

| CAT. CODE • | ENTITY BEING MAINTAINED OR CHANGED |
|---|---|
| **3000** | **Knowledge, Technology, and Art Form Outcomes,** *continued* |
| 3400 | Art Forms and Works |
| 3410 | Architecture |
| 3420 | Dance |
| 3430 | Debate and Oratory |
| 3440 | Drama |
| 3450 | Literature and Writing |
| 3460 | Music |
| 3470 | Painting, Drawing, and Photography |
| 3480 | Sculpture |
| 3490 | Other Fine Arts |
| 3500 | Other Knowledge, Technology, and Art Form Outcomes |
| **4000** | **Resource and Service Provision Outcomes** |
| 4100 | Provision of Facilities and Events |
| 4110 | Provision of Facilities |
| 4120 | Provision of Sponsorship of Events |
| 4200 | Provision of Direct Services |
| 4210 | Teaching |
| 4220 | Advisory and Analytic Assistance |
| 4230 | Treatment, Care, and Referral Services |
| 4240 | Provision of Other Services |
| 4300 | Other Resource and Service Provision Outcomes |
| **5000** | **Other Maintenance and Change Outcomes** |
| 5100 | Aesthetic-Cultural Activities, Traditions, and Conditions |
| 5200 | Organizational Format, Activity, and Operation |
| 5300 | Other Maintenance and Change |

## Appendix B. Basic Skills Tests for College Students

**Nelson-Denny Reading Test, Forms E and F**

Purpose: To evaluate reading comprehension, vocabulary development, and reading rate.

Use: To screen, predict college success, and to diagnose reading difficulties.

Reliability: Test-retest reliabilities for vocabulary subtest is .89 to .95, for comprehension subtest is .75 to .82, for reading rate is .62 to.82.

Validity: Limited information available; comprehension subtest is context dependent.

**Degrees of Reading Power**

Purpose: To measure reading effectiveness.

Use: To predict probability of success for students in prose materials of varying difficulties.

Reliability: K-R 20 coefficients vary between .93 and .97.

Validity: Correlations with CAT reading test range from .77 to .85.

**Prescriptive Reading Performance Test**

Purpose: To evaluate reading and spelling patterns and determine how a student employs visual and auditory modalities in reading.

Use: To assess preliminarily reading level and reading comprehension; to identify strengths and weaknesses.

Reliability: Test-retest is .98, Spearman-Brown corrected split-half is .98.

Validity: Pearson correlations with six other reading measures range from .65 to .94.

**Self-scoring Reading Placement Test**

Purpose: To evaluate reading and mathematical skills necessary for success in a two-year college for students entering postsecondary institutions with open-door policies.

Use: To assist in placing students in college courses.

Reliability: K-R 20 coefficients available in manual.

Validity: Predictive validity of English and mathematics grades range from .06 to.70, with a median of .40.

## Appendix B *(continued)*

**Test of Mathematical Abilities**

Purpose: To assess mathematical attitudes and aptitudes.

Use: For individual assessment.

Reliability: Internal consistency reliabilities range from .96 to .57.

Validity: Correlations of .26 to .31 between attitude subscale scores and three standardized mathematics tests.

**AH Vocabulary Scale**

Purpose: To evaluate vocabulary level.

Reliability: K-R 21 coefficients range from .60 to .90, split-half coefficients range from .70 to .90 for all 80-word tests.

Validity: Correlates between .50 and .75 with vocabulary tests and variables such as non-verbal, reading, and mathematics ability and intelligence.

**Comprehensive Tests of Basic Skills: Reading**

Purpose: To evaluate reading and reference skills.

Reliability: K-R 20 coefficients for total reading scores are .94 to .97, for reference skills are .76 to .94.

Validity: Information not available.

**Iowa Silent Reading Tests**

Purpose: To measure vocabulary, reading comprehension, and reading efficiency.

Reliability: Median alternate-forms reliabilities for vocabulary, comprehension, and efficiency are .86, .83, and .77, respectively.

Validity: No predictive validity information is available, correlations with other reading tests are in the .70s and .80s.

**Reading Progress Scale**

Purpose: To evaluate "reading-input" performance.

Reliability: Alternate-form estimate is .84.

Validity: Only available for grades 3-6.

## Appendix B *(continued)*

**Sequential Tests of Educational Progress: Reading Series II**

Purpose: To measure sentence and passage comprehension.

Reliability: Internal consistency and alternate forms reliability information are available in manual.

**College Board Achievement Test in Mathematics, Level 1**

Purpose: To measure mathematical abilities of persons with at least three years of college preparatory mathematics courses.

Reliability: K-R 20 coefficient is .88.

Validity: Good predictor of college grades but does not add much predictive ability when SAT scores and high school GPA are already included.

*Information for this appendix was gathered from reviews in the *Mental Measurement Yearbook*

**References**

[NOTE: In some instances references have been made to early drafts of works that subsequently may have been published in another form.—J.A. and J.S.]


Aiken, L. R. (1985). Review of ACT assessment program. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 29-31). Lincoln, NE: University of Nebraska Press.

Aleamoni, L. M. (1978). Review of CLEP general examinations. In O. T. Buros (Ed.), *The eighth mental measurements* (pp. 29-30). Highland Park, NJ: The Gryphon Press.

Association of American Colleges. (1985, February). *Integrity in the college curriculum: A report to academic community*. Washington, DC: Association of American Colleges.

Astin, A. (1963). Further validation of the Environmental Assessment Technique. *Journal of Educational Psychology, 54*, 217-226.

Astin, A. (1968). Undergraduate achievement and institutional excellence. *Science, 161*, 661-668.

Astin, A. W. (1974). Measuring the outcomes of higher education. In H. R. Bowen (Ed.), Evaluating institutions for accountability, *New directions for institutional research* (pp. 23-46). San Francisco: Jossey-Bass.

Astin, A. W. (1977). *Four critical years*. San Francisco: Jossey-Bass.

Astin, A. W. (1980). When does a college deserve to be called 'high quality?'. *Current Issues in Higher Education, 1*, 1-9.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297-308.

Astin, A. W. (1985). *Achieving educational excellence*. San Francisco: Jossey-Bass.

Astin, A., & Panos, R. (1969). *The educational and vocational development of college students*. Washington, DC: American Council on Education.

Ayres, Q. (1983). Student achievement at predominantly white and predominantly black universities. *American Educational Research Journal, 20*, 291-304.

Ayres, Q., & Bennett, R. (1983). University characteristics and student achievement. *Journal of Higher Education, 54*, 516-532.

Bean, J. P., & Kuh, G. D. (1984). The reciprocity between student-faculty informal contact and academic performance of university undergraduate students. *Research in Higher Education, 21*, 461-477.

Bennett, W. J. (1984). *To reclaim a legacy: A report on the humanities in higher education*. Washington, DC: National Endowment for the Humanities.

Bowen, H. R. (1974). The products of higher education. In H. R. Bowen (Ed.), Evaluating institutions for accountability. *New directions for institutional research* (pp. 1-22). San Francisco: Jossey-Bass.

Bowen, H. R. (1977). *Investment in learning: The individual and social value of American higher education*. San Francisco: Jossey-Bass.

Bowen, H. R. (1979). Evaluating educational quality: A conference summary. In A. W. Astin (Ed.), *Evaluating educational quality: A conference summary*. Washington, DC: Council on Postsecondary Accreditation.

Brabeck, M. (forthcoming). Critical thinking skills and reflective judgment development: Redefining the aims of higher education. *Journal of Applied Developmental Psychology*.

Brown, R. D. (1985). Review of college student experiences. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 365-366). Lincoln, NE: University of Nebraska Press.

Buros, O. T. (Ed.). (1978). *The eighth mental measurements yearbook*. Highland Park, NJ: The Gryphon Press.

Centra, J., & Rock, D. (1971). College environments and student academic achievement. *American Educational Research Journal, 8*, 623-634.

Cohn, S. J. (1985). Review of Graduate Record Examinations – General Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp.622-624). Lincoln, NE: University of Nebraska Press.

Dumont, R., & Troelstrup, R. (1981). Measures and predictors of educational growth with four years of college. *Research in Higher Education, 14*, 31-27.

Educational Testing Service. (1976). *Undergraduate Assessment Program Guide*. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (1978). *Undergraduate Assessment Program Guide*. Princeton, NJ: Educational Testing Service.

Edwards, J. E., & Waters, L. K. (1983). Predicting university attrition: A replication and extension. *Educational and Psychological Measurement, 43*, 233-236.

Eisert, D., & Tomlinson-Keasey, C. (1978). Cognitive and interpersonal growth during the college freshman year: A structural analysis. *Perceptual and Motor Skills, 46*, 995-1005.

Ewell, P. T. (1983). *Information on student outcomes: How to get it and how to use it*. Boulder, CO: NCHEMS.

Ewell, P. T. (1984). *The self-regarding institution: Information for excellence*. Boulder, CO: NCHEMS.

Ewell, P. (1985). Assessment: What's it all about? *Change, 17*(6), 32-36.

Feldman, K. A., & Newcomb, T. M. (1969). *The impact of college on students*. San Francisco: Jossey-Bass.

Forrest, A. F., and Steele, J. M. (1978). *College Outcomes Measures Project*. Iowa City, IA: American College Testing Program.

Grant, M. K., & Hoeber, D. R. (1978). *Basic skills programs: Are they working?* Washington, DC: American Association for Higher Education.

Harris, J. (1985). *Assessing outcomes in higher education: Practical suggestions for getting started*. Washington, DC: American Association for Higher Education.

Hartle, T. W. (1985). *The growing interest in measuring the educational achievement of college students*. Washington, DC: American Association for Higher Education.

Harvey, P., & Lannholm, G. (1960). *Achievement in three major fields during the last two years of college*. (Graduate Record Examination Special Report No. 60-Z). Princeton, NJ: Educational Testing Service.

Hills, J. R. (1978). Review of ACT Assessment. In O. T. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 622-626). Highland Park, NJ: The Gryphon Press.

Jaegar, R. M. (1985). Review of Graduate Record Examinations General Test. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 624-626). Lincoln, NE: University of Nebraska Press.

Keeley, S., Browne, M., & Kreutzer, J. (1982). A comparison of freshmen and seniors on general and specific essay tests of critical thinking. *Research in Higher Education, 17*, 139-154.

Kifer, E. (1985). Review of ACT Assessment Program. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 31-36). Lincoln, NE: University of Nebraska Press.

Lehmann, I. (1963). Changes in critical thinking, attitudes and values from freshman to senior years. *Journal of Educational Psychology, 54*, 305-315.

Lenning, O. T., Lee, Y. S., Micek, S. S., & Service, A. L. (1977). *A structure for the outcomes of postsecondary education.* Boulder, CO: NCHEMS.

Lenning, O. T., Munday, L., & Maxey, J. (1969). Student educational growth during the first two years of college. *College and University, 44,* 145-153.

McKeachie, W. J., Pintrich, P. R., Lin, Y., & Smith, D. A. F. (1986). *Teaching and learning in the college classroom: A review of the research literature.* (Tech. Rep. No. 86-13-001.0.) Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Mentkowski, M., & Doherty, A. (1984). Abilities that last a lifetime. *AAHE Bulletin, 36*(6), 3-14.

Mentkowski, M., & Strait, M. J. (1983). *A longitudinal study of student change in cognitive development, learning strategies, and generic abilities in an outcome-centered liberal arts curriculum.* (Final Report to the National Institute of Education: Research report No. 6). Milwaukee, WI: Alverno College Productions. (ERIC Document Reproduction Service No. ED 239 562).

Merwin, J. C. (1978). Review of National Teacher Examinations. In O. T. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 514-516). Highland Park, NJ: The Gryphon Press.

Milholland, J. E. (1978). Review of CLEP General Examinations. In O. T. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 30-32). Highland Park, NJ: The Gryphon Press.

Miller, J. K. (1985). Review of college student experiences. In J. V. Mitchell (Ed.), *The ninth mental measurements yearbook* (pp. 365-366). Lincoln, NE: University of Nebraska Press.

Mitchell, J. V. (1978). Review of National Teacher Examination. In O. T. Buros (Ed.), *The eighth mental measurements yearbook* (pp. 516-518). Highland Park, NJ: The Gryphon Press.

Mitchell, J. V. (Ed.). (1985). *The ninth mental measurements yearbook.* Lincoln, NE: University of Nebraska Press.

Munro, B. H. (1981). Dropouts from higher education: Path analysis of a national sample. *American Educational Research Journal, 16,* 133-141.

Nichols, R. (1964). Effects of various college characteristics on student aptitude test scores. *Journal of Educational Psychology, 55,* 45-54.

Pace, C. R. (1979). *Measuring outcomes of college: Fifty years of findings and recommendations for the future.* San Francisco: Jossey-Bass.

Pace, C. R. (1984). *Measuring the quality of college student experiences.* Los Angeles: University of California at Los Angeles, Higher Education Research Institute.

Pascarella, E. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research, 50,* 545-595.

Pascarella, E. (1986). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1, pp.1-61). New York: Agathon Press.

Pascarella, E., Duby, P. B., & Iverson, B. K. (1983). A test and reconceptualization of a theoretical model of college withdrawal in a commuter institution setting. *Sociology of Education, 56*(2), 88-100.

Pascarella, E. T., Smart, J. C., & Ethington, C. A. (1986). Long-term persistence of two-year college students. *Research in Higher Education, 24,* 47-71.

Perry, W. (1970). *Forms of intellectual and ethical development in the college years.* New York: Holt, Rinehart and Winston.

Rock, D., Centra, J., & Linn, R. (1970). Relationships between college characteristics and student achievement. *American Educational Research Journal, 7,* 109-121.

Schmidt, J., & Davison, M. (1981). Reflective judgment: How to tackle the tough questions. *Moral Education Forum, 6,* 2-14.

Southern Regional Education Board. (1985). Access to quality undergraduate education. *The Chronicle of Higher Education*, July 3, 9-12.

Stewart, A. (1977). *Analysis of argument: An empirically derived measure of intellectual flexibility*. Boston: McBer.

Study group on the conditions of excellence in American higher education. (1984). *Involvement in learning: Realizing the potential of American higher education*. Washington, DC: National Institute of Education.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*, 89-125.

Turnbull, W. W. (1985). Are they learning anything in college? *Change, 17*(6), 23-26.

Volkwein, J. F., King, M. C., & Terenzini, P. T. (1986). Student-faculty relationships and intellectual growth among transfer students. *The Journal of Higher Education, 57*, 413-430.

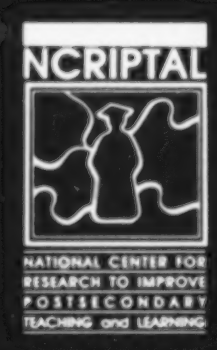
Watson, G., & Glaser, E. (1964). *Critical thinking appraisal*. New York: Harcourt Brace Jonavich.

Winteler, A. (1981). The academic department as environment for teaching and learning. *Higher Education, 10*, 25-35.

Winter, D., & McClelland, D. (1978). Thematic analysis: An empirically derived measure of the effects of liberal arts education. *Journal of Educational Psychology, 70*, 8-16.

Winter, D., McClelland, D., & Stewart, A. (1981). *A new case for the liberal arts: Assessing institutional goals and student development*. San Francisco: Jossey-Bass.

Wolfe, L. (1983). Effects of higher education on ability for blacks and whites. *Research on Higher Education, 19*, 3-10.


NCRIPTAL
NATIONAL CENTER FOR
RESEARCH TO IMPROVE
POSTSECONDARY
TEACHING and LEARNING